



A stereo range computation method using thermal infrared and visible cameras



Wen Chen, Li Cao^{*}, Zijian Wu, Shengguo Huang

College of Civil Aviation, Nanjing University of Aeronautics & Astronautics, Nanjing, Jiangsu, PR China

HIGHLIGHTS

- A multimodal stereo ranging system composed of a thermal camera and a visible camera is built.
- A simple and reliable range computation method is presented when two different cameras are placed in parallel configuration.
- The procedure to identify the parameters required for the range computation method is described.
- The effect of the relative position of a certain target and reference points is also analyzed when only two reference points are used to identify the parameters.

ARTICLE INFO

Article history:

Received 16 April 2013

Available online 19 November 2013

Keywords:

Binocular stereo

Range computation

Thermal camera

Visible camera

Parallel camera configuration

ABSTRACT

Binocular stereo vision can provide geometric position information of the target, which enables one to track the moving object precisely. Although visible images are full of details of geometry and texture, it is difficult to detect moving objects in poor visibility. Appropriate fusion of infrared and visible images can combine the complementary information and obtain a better description of the scene, which will help in target detection and target localization. Considering the physical differences between thermal infrared cameras and visible cameras, a stereo ranging method in parallel camera configuration is proposed in this paper. The rotation center of a pan-tilt device is used as the origin of the coordinate system in the method. It could not only locate the object in poor visibility, but also determine the parameters essential in range computation in practical work, which is very difficult in traditional calibration methods. Furthermore, the effect of the relative position of a certain target and reference points is analyzed. Experiments also proved the validity of the proposed method.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Monocular vision is simple because it could avoid image fusion, which means it can meet the requirement of real-time in visual surveillance. However, it can only provide two-dimensional information of the scene. On the other side, binocular vision can provide three-dimensional (3D) information of the object by shooting the object at the same time by two cameras placed in different positions. Thus, binocular vision makes it possible to precisely track the moving object. However, it is a huge challenge to detect and track the moving object by two visible cameras in poor visibility, such as night vision, fog, rain, and other inclement weather.

With the rapid development of science and technology, the application of multi-sensor technology is increasingly widespread. Multiple vision sensors are now widely used in visual sur-

veillance, improving the robustness and accuracy of visual surveillance system in all-weather conditions. A target ranging system based on visual and active infrared imaging in night vision was proposed in [1]. An active infrared detection system is usually consisted of an active infrared transmitter and a passive infrared sensor (receiver). The transmitter emits a beam of light into the detection zone. The light, which is reflected by the background, returns to the receiver that constantly monitors the detection zone. Therefore, the effective range of an active infrared system is dependent on the power of the transmitter. Furthermore, the transmitter and its power supply are not convenient in field. By contrast, a passive infrared detection system can work without transmitters. Only one thermal camera is needed in the passive system. The thermal camera creates images based on differences in surface temperature by detecting infrared radiation (heat) that emanates from objects and their surrounding environment. It can be used in imaging when there is insufficient visible light to see such as in night vision, as well as special

^{*} Corresponding author. Tel.: +86 13601585169.

E-mail address: caoli@nuaa.edu.cn (L. Cao).

camouflage detection. However, the thermal image is poor in spatial resolution and texture which means a texture or an edge in a visible image is often missing in the thermal image. Using visible-infrared camera pairs is now attracting more attention because the combination of the information from these two sensors performs well in complementary situations. Although related work has been developed in visual-thermal fusion, most of them are focused on enhanced vision [2–4]. A multimodal stereo vision system consisted of a thermal camera and a visible camera is able to make full use of grayscale and temperature to get the depth information of the object in poor visibility. Therefore, it contributes to precise tracking in all weather conditions [5,6].

The intrinsic parameters of the lens (focal length and principle point), the extrinsic parameters of the lens (relative position of the optical centers), and the disparity information are indispensable in stereo ranging. Therefore, the precision of ranging results greatly depends on the precision of the parameters calibrated. In a typical stereo vision system composed of two visible cameras, the calibration is usually carried out with the help of a certain calibration board such as a checkerboard pattern. With a matter of placing a checkerboard pattern in front of the camera, these parameters can be carefully calibrated. However, most of the existing calibration methods are complicated and time-consuming [7–12]. It usually costs several hours to calibrate the parameters for just one camera. Furthermore, once the focal length changes or one of the cameras moves, it is necessary to re-calibrate the intrinsic/extrinsic parameters. So they are not appropriate in field.

It is obvious that calibrating these parameters is in the first place to compute the range in the multimodal stereo vision system composed of a thermal camera and a visible camera of which intrinsic parameters are different. Meanwhile, the calibration method must be simple and reliable to be suitable for practical application. Furthermore, a thermal image is very different from a visible image. If the calibration method based on the calibration board is adopted, some extra care must be taken to ensure the calibration board looks similar in each modality. In [13], the multimodal stereo calibration was carried out with a heated calibration board based on the typical calibration method. As mentioned above, although the calibration results can meet the requirement of precision, the procedure is too complicated to be applied in field. In [14], the thermal camera was calibrated with the help of a pair of calibrated visible cameras. However, it required ideal environmental conditions.

In this paper, the range is computed to provide position information of the target for the tracking system, so operation speed is more important than precision. To solve these problems, a stereo ranging method in parallel camera configuration is proposed in this paper. The rotation center of a pan-tilt device is used as the origin of the coordinate system in the method. Based on this method, the line distance between the target and the rotation center of the pan-tilt device can be computed as well as the angle. The procedure is also shown to identify the parameters used in this method. The method is simple and easy to calculate. Only geometry information of the pan-tilt devices and several reference points (at least two points) are needed to determine the parameters in the proposed method. The effect of the relative position of a certain target and reference points is also analyzed when only two reference points are used to identify the parameters. The region in which better ranging result can be gotten is provided in the experiment. Finally, experiments proved that the results of the proposed method could meet the requirement of object tracking and locating in all weather conditions.

2. Computing range

In stereo vision, the relationship between a 3D point and its image projection is determined by the camera model. The simplest model of a camera is the pinhole camera model. In this paper, pinhole model is used as the camera model for both visible camera and thermal camera. In addition, the position of the image plane and the position of the optical center are switched in the coordinate system described in this paper for convenience (see Fig. 1). As shown in Fig. 1, l is the image plane, O is the optical center of the lens, $O_o(u_o, v_o)$ is the intersection of the optical axis and the image plane (the principle point), $p(u, v)$ is the projection of the spatial point $P(x_c, y_c, z_c)$ to image coordinates, and f is the focal length, which is precisely the distance from the optical center to the image plane for a idealized pinhole camera. Usually, the optical center O is set as the origin of the camera coordinate system (CCS), the optical axis is set as Z-axis, and the optical axis is perpendicular to the image plane. Based on this assumption, the relationship between a 3D point P and its image projection p is given by:

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & u_o \\ 0 & f & v_o \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix}. \quad (1)$$

2.1. Stereo ranging principle in parallel camera configuration

According to the relationship of the optical axes of the right and left lenses, there are two methods of binocular stereoscopic camera with controllable vergence: toed-in configuration and parallel configuration. In toed-in configuration, these optical axes are made to cross at some point during the shooting; while in parallel configuration, the optical axes are parallel to each other. The latter is relatively easier in computing range so the calibration is easier. Therefore, this study computed range using a visible camera and a thermal camera which were arranged in parallel configuration.

An example of parallel camera configuration is shown in Fig. 2(a). The camera coordinate system of the left camera C_1 is $O_1x_1y_1z_1$, and the camera coordinate system of the right camera C_2 is $O_2x_2y_2z_2$. Let P be the target point, then p_1 and p_2 are the projections of P to the image plane in the left and right camera view respectively. The intersection of O_1p_1 and O_2p_2 is P , and it is the one which is only confirmed. In the parallel configuration, each axis in $O_1x_1y_1z_1$ is parallel to the correspondence axis in $O_2x_2y_2z_2$. The distance from O_1 to O_2 in X-axis is b , the interval of O_1 and O_2 in Y-axis is e , and the interval of O_1 and O_2 in Z-axis is d (when O_1 is above O_2 , $e > 0$; when O_2 is behind O_1 , $d > 0$). In the configuration described above, set the left camera coordinate system

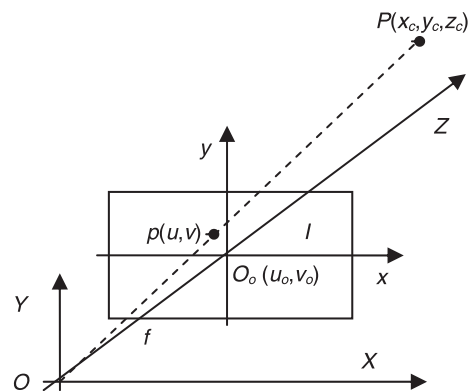


Fig. 1. Projection model.

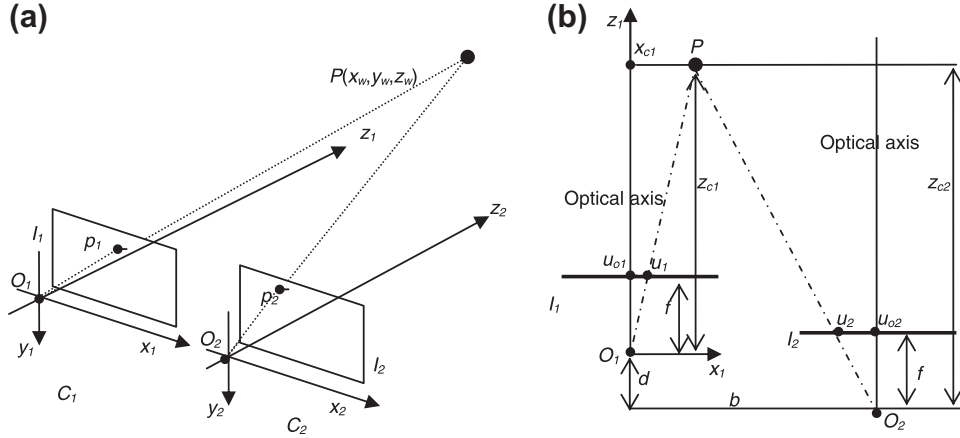


Fig. 2. Parallel arrangement of the two optical axes. (a) 3D view of the parallel configuration and (b) X-Z plane projection of the geometry of parallel cameras.

$O_1x_1y_1z_1$ as the world coordinate system (WCS) and the relationship between WCS and CCS is shown as follows:

$$\begin{cases} x_w = x_{c1} = x_{c2} + b \\ y_w = y_{c1} = y_{c2} + e \\ z_w = z_{c1} = z_{c2} - d \end{cases} \quad (2)$$

In order to straightforwardly describe the geometric relationship in the parallel configuration, 3D view of the configuration shown in Fig. 2(a) is converted to 2D view. Fig. 2(b) shows X-Z plane projection of the geometry of parallel cameras. Suppose that the focal length of each camera is f , the principal point of C_1 is (u_{o1}, v_{o1}) and the principal point of C_2 is (u_{o2}, v_{o2}) . The coordinate of P in the left camera coordinate system is (x_{c1}, y_{c1}, z_{c1}) and the coordinate of P in the right camera coordinate system is (x_{c2}, y_{c2}, z_{c2}) . (u_1, v_1) is the coordinate of p_1 on the 2-D image plane I_1 and (u_2, v_2) is the coordinate of p_2 on the 2-D image plane I_2 . In the configuration described above, the coordinates centered on O_1 of the left camera is selected as the coordinates in which the calculations will be done. In Fig. 2(b), the relationship between the image coordinates and the camera coordinates can be seen by similar triangles that:

$$u_1 - u_{o1} = f \frac{x_{c1}}{z_{c1}} \quad (3)$$

$$u_2 - u_{o2} = f \frac{x_{c1} - b}{z_{c1} + d} \quad (4)$$

$$v_1 - v_{o1} = f \frac{y_{c1}}{z_{c1}} \quad (5)$$

$$v_2 - v_{o2} = f \frac{y_{c1} - e}{z_{c1} + d} \quad (6)$$

Thus, x_{c1} , y_{c1} and z_{c1} can be solved by these equations:

$$y_{c1} = \frac{(v_1 - v_{o1})[bf + d(u_2 - u_{o2})]}{f(u_1 - u_2 + u_{o2} - u_{o1})} \quad (7)$$

$$x_{c1} = \frac{(u_1 - u_{o1})[bf + d(u_2 - u_{o2})]}{f(u_1 - u_2 + u_{o2} - u_{o1})} \quad (8)$$

$$z_{c1} = \frac{bf + d(u_2 - u_{o2})}{(u_1 - u_2 + u_{o2} - u_{o1})} \quad (9)$$

The expression of z_{c1} shows that it can be solved by Eqs. (3) and (4) which are related to X-coordinates and Z-coordinates. In this paper, the range is computed to get the position information of the tracking

object, so it is not essential to obtain the information in Y-axis especially when the object only moves on the ground. Furthermore, the computational complexity can be reduced if the information in Y-axis is ignored. Therefore, only information in X-axis and Z-axis is considered in stereo ranging in our paper. The above is the stereo ranging principle in parallel camera configuration.

2.2. Proposed ranging equations

Eqs. (8) and (9) show that z_{c1} is a function of f , u_{o1} , u_{o2} , u_1 , u_2 , b and d . The focal length f of visible camera and thermal camera can be obtained from each instruction and it is not difficult to choose the lenses with the same focal length. u_1 and u_2 can be observed from images or obtained through stereo matching. u_{o1} and u_{o2} is the x-coordinate of the principal point of each camera, and they should be on the center of the image plane in ideal conditions. b and d describe the geometric relationship of O_1 and O_2 , which means they are the extrinsic parameters. However, the optical center is not a well-defined, constant, physical location. It actually moves as the lens focuses and zooms, making it difficult physically to measure b and d . Most of the stereo ranging systems obtained b and d by an off-line calibration [9]. As mentioned above, once the focal length changes or the camera moves, it is necessary to recalibrate the extrinsic parameters. Furthermore, the classic calibration procedure is complicated and time-consuming.

The location of optical center is invisible and uncertain, which makes the measurement of b and d difficult. In contrast, the camera rotation center is a well-defined location, around which the pan-tilt device rotates the cameras. Therefore, the distance between the two rotation centers is easier to measure in practical use. Instead of O_1 , the optical center of the left camera, the left rotation center R_1 is selected as the origin of the coordinate system in this paper. Thus, every geometric relationship related to the optical centers can be transformed to the geometric relationship related to the rotation centers. Accordingly, the distance from P to R_1 rather than O_1 is the distance to be solved in this paper.

In the presented method, two cameras are separately mounted on two pan-tilt devices with the same configuration. R_1 , the rotation center of the left camera is set as the origin of XR_1Z . As shown in Fig. 3, the rotation center R_i does not coincide with the optical center O_i ($i = 1, 2$). The parameter with the subscript i refers to the left camera when $i = 1$ and refers to the right camera when $i = 2$, similarly hereinafter. In X-Z plane, the distance between R_i and O_i is m_i in X-axis and n_i in Z-axis. The distance between R_1 and R_2 is r in X-axis and t in Z-axis. In XR_1Z , when O_i is to the right of R_i , $m_i > 0$; when O_i is behind R_i , $n_i > 0$; when O_2 is behind O_1 , $t > 0$. x_0

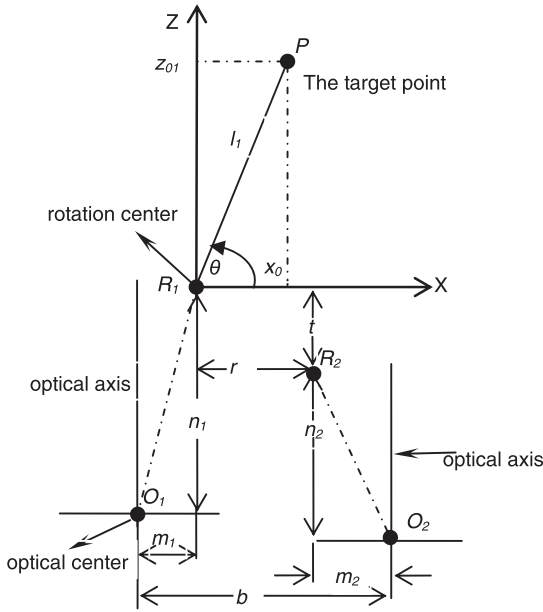


Fig. 3. The geometric relationship of the rotation centers and the optical centers.

is the distance from P to R_1 in X -axis and z_{01} is the distance from P to R_1 in Z -axis. In the configuration described above, when the optical axes are parallel the relationship between b and r and the relationship between d and t can be described as follows:

$$b = r + m_2 - m_1 \quad (10)$$

$$d = t + n_2 - n_1. \quad (11)$$

Accordingly, the relationship between x_0 and x_{c1} and the relationship between z_{01} and z_{c1} can be described as follows:

$$x_{c1} = x_0 - m_1 \quad (12)$$

$$z_{c1} = z_{01} + n_1. \quad (13)$$

Therefore, in XR_1Z , Eqs. (3) and (4) are transformed as follows:

$$u_1 - u_{o1} = \frac{f(x_0 - m_1)}{z_{01} + n_1} \quad (14)$$

$$u_2 - u_{o2} = \frac{f(x_0 - r - m_2)}{z_{01} + t + n_2}. \quad (15)$$

Thus, x_0 and z_{01} can be solved from these equations:

$$x_0 = \frac{(u_1 - u_{o1})[f(r + m_2 - m_1) + (u_2 - u_{o2})(t + n_2 - n_1)]}{f(u_1 - u_2 + u_{o2} - u_{o1})} + m_1 \quad (16)$$

$$z_{01} = \frac{f(r + m_2 - m_1) + (u_2 - u_{o2})(t + n_2 - n_1)}{(u_1 - u_2 + u_{o2} - u_{o1})} - n_1. \quad (17)$$

Further, the line distance from P to R_1 and angle between the connecting line of P and R_1 and X -axis can be obtained from Eqs. (18) and (19).

$$l_1 = \sqrt{x_0^2 + z_{01}^2} \quad (18)$$

$$\theta = \arctan\left(\frac{z_{01}}{x_0}\right) \quad (19)$$

Finally, x_0 and z_{01} can be figured out when f , u_{oi} , u_i , r , t , m_i and n_i are already known. Then, l_1 and θ can be obtained too.

2.3. Parameter identification

The parameters in Eqs. (16) and (17) such as f , u_{oi} , u_i , r and t can be obtained through the way mentioned in last section. However, m_i and n_i are still unknown. These four parameters must be identified to figure out x_0 and z_{01} . According to Eqs. (16) and (17), x_0 and z_{01} can be solved when all of the other parameters are already known. Alternatively, m_i and n_i can be solved when all of the other parameters are known too.

For a fixed ranging system in which f , u_{oi} , r and t are known and constant, a point of which coordinates on two image planes and coordinates in XR_1Z are already known can provide two equations, and two known points can provide four equations. Thus a minimum of two points are needed to solve for the four parameters.

Suppose there are two spatial reference points Q_1 and Q_2 and their coordinates on two image planes and coordinates in XR_1Z are already known. $x_0(1)$ is the distance from Q_1 to R_1 in X -axis and $z_{01}(1)$ is the distance from Q_1 to R_1 in Z -axis. $u_1(1)$ is the x -coordinate of Q_1 on the left image plane and $u_2(1)$ is the x -coordinate of Q_1 on the right image plane. $x_0(2)$ is the distance from Q_2 to R_1 in X -axis and $z_{01}(2)$ is the distance from Q_2 to R_1 in Z -axis. $u_1(2)$ is the x -coordinate of Q_2 on the left image plane and $u_2(2)$ is the x -coordinate of Q_2 on the right image plane. Thus the following equations can be obtained:

$$\begin{cases} u_1(1) - u_{o1} = \frac{f(x_0(1) - m_1)}{z_{01}(1) + n_1} \\ u_2(1) - u_{o2} = \frac{f(x_0(1) - r - m_2)}{z_{01}(1) + t + n_2} \\ u_1(2) - u_{o1} = \frac{f(x_0(2) - m_1)}{z_{01}(2) + n_1} \\ u_2(2) - u_{o2} = \frac{f(x_0(2) - r - m_2)}{z_{01}(2) + t + n_2} \end{cases} \quad (20)$$

m_1 , m_2 , n_1 , and n_2 can be figured out by (20).

$$\begin{cases} m_1 = \frac{(z_{01}(1) - z_{01}(2))(u_1(1) - u_{o1})(u_1(2) - u_{o1}) - f(u_1(2) - u_{o1})x_0(1) + f(u_1(1) - u_{o1})x_0(2)}{f(u_1(1) - u_1(2))} \\ m_2 = \frac{(z_{01}(1) - z_{01}(2))(u_2(1) - u_{o2})(u_2(2) - u_{o2}) - f(u_2(2) - u_{o2})x_0(1) + f(u_2(1) - u_{o2})x_0(2)}{f(u_2(1) - u_2(2))} - r \\ n_1 = \frac{u_{o1}(z_{01}(1) - z_{01}(2)) + f(x_0(1) - x_0(2)) + u_1(2)z_{01}(2) - u_1(1)z_{01}(1)}{u_1(1) - u_1(2)} \\ n_2 = \frac{u_{o2}(z_{01}(1) - z_{01}(2)) + f(x_0(1) - x_0(2)) + u_2(2)z_{01}(2) - u_2(1)z_{01}(1)}{u_2(1) - u_2(2)} - t \end{cases} \quad (21)$$

(21) shows unique solution when the cameras are fixed. In practical operation, r , t can be measured directly and x_{o1} , z_{o1} will be gotten by simple geometric operations (See Appendix).

In sum, for a fixed ranging system in which f , u_{oi} , r and t are known and constant, only two reference points with the parameters l_1 , l_2 , u_1 and u_2 known are needed to solve for m_i and n_i . With the identified parameters, the line distance from P to R_1 can be figured out as well as the corresponding angle θ . The measurement of the parameters involved such as r , t , l_1 and l_2 is relatively easy in operation. Other than the traditional calibration method, the parameters in this paper can be identified without certain calibration board. One could identify the essential parameters by arbitrarily selecting two points in the field of view common to the cameras as the reference points and measuring the line distances between each point and each rotation center. And the calculation is simple.

3. Experimental results

In this section, experimental results are presented for the range computation. Two cameras are mounted on a pair of pan-tilt devices and they can translate horizontally and rotate left-right. Thus, the optical axes can be set in parallel configuration by rotating the pan-tilt devices. As mentioned in Section 2.3, the parameters can be identified through the steps shown in Fig. 4. As ranging is the purpose in this paper, the match of

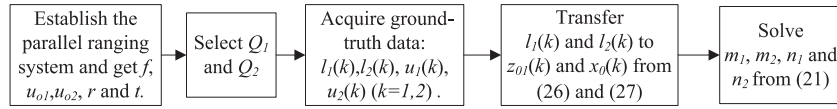


Fig. 4. The parameter identification steps.

conjugate points in the stereo pair of images is not discussed here. $u_1(k)$ and $u_2(k)$ ($k = 1, 2$) are matched manually in this experiment. With the identified parameters, l_1 and θ of a certain target can be figured out.

In 3.1, the effect of the relative position of the target and reference points on ranging accuracy is analyzed when only two reference points are used to identify the parameters. In 3.2, the range from a moving person to R_1 is computed with the proposed ranging method.

3.1. The relative position of the target and two reference points

Although m_1, m_2, n_1 , and n_2 can be determined once the cameras are fixed, there are still some factors that affect the ranging accuracy. In this part, we will discuss the effect of the relative position of the target and reference points.

In order to facilitate analysis, we divide the scene in front of two cameras into four regions, **RA**, **RB**, **RC** and **RD**. As shown in Fig. 5(a), Q_1 and Q_2 are reference points, R_1 and R_2 are rotation centers of two cameras. Fig. 5(b) shows the experimental scheme. Objects, A_1 – A_7 and T_1 – T_3 , are arranged in two rows. Considering the halation in thermal images, the center of each object is chosen as the spatial point in this experiment. Different combinations of reference points are also shown in Fig. 5(b). The information of these spatial points is listed in Table 1.

Fig. 6 illustrates the real images captured by the cameras. The left one is the thermal image and the right one is the visible image. The water temperature in the cups was about 80 °C and the room temperature was about 20 °C. The focal length of the thermal camera is 395 pixel. Suppose that the principal point is on the center of the image, $u_{o1} = 160$. The focal length of the visible camera is also set to 395 pixel and $u_{o2} = 158$. By measuring, parameters r and t could be gotten: $r = 133$ mm, $t = 12$ mm (the left camera was behind the right). Therefore, the parameters can be identified by the steps described before and the range of a target can be computed. The ranging accuracy is quantified in terms of $|E_x|$ and $|E_z|$, calculated by (22) and (23).

$$E_x = x_{0c} - x_{0t} \quad (22)$$

$$E_z = z_{01c} - z_{01t} \quad (23)$$

Here z_{01t} and x_{0t} are the ground-truth data, and z_{01c} and x_{0c} are the computed results. In this paper, $|E_x|$ and $|E_z|$ are measured in millimeter.

When T_1, T_2 and T_3 are selected as the targets and two of A_1 – A_7 are selected as reference points, the target is either in region **RB** or in **RC**, the dashed lines shown in Fig. 5(b). When A_2 and A_5 are selected as the targets and two of T_1 – T_3 are selected as reference points, the target is either in region **RA** or in **RD**, the solid lines in Fig. 5(b). A target, to which region it will belong, **RA**, **RB**, **RC** or **RD**, depends on (Q_1, Q_2) . For example, T_2 is in **RC** when $(Q_1, Q_2) = (A_1, A_2)$ and it is in **RB** when $(Q_1, Q_2) = (A_4, A_5)$. Table 2 lists the ranging results of different targets computed with different pairs of reference points, (Q_1, Q_2) . A comparison can be made based on the results.

From Table 2, one can find that $|E_x|$ will be smaller when a certain target is in **RB** or **RA** than in **RC** or **RD** in general. There is a similar result when different targets are computed with a certain pair of reference points (Q_1, Q_2) .

For a certain pair of reference points (Q_1, Q_2) , $|E_z|$ is constant for different targets because the targets own the same value of z_{01t} . For a certain target, $|E_z|$ is smaller when Q_1 is closer to the principal point.

Therefore, in order to get better ranging result of a certain target based on two reference points, selecting Q_1 and Q_2 to insure the target being in region **RA** and selecting the point close to the principal point in the image plane as Q_1 will be better. It is also verified in the next experiment run in practice.

Fig. 7 illustrates the real images captured by the parallel ranging system. The hot water in the bottle and the energized adapter are selected as the heat sources. The related parameters are shown as follows: $r = 155$ mm, $t = -120$ mm (the left was behind the right), $f = 395$ pixel, $u_{o1} = 160$, $u_{o2} = 178$. Q_1 and Q_2 are shown in Fig. 7. Their information is listed in Table 3. Thus the parameters can be figured out and the results are shown as follows:

$$m_1 = -27.6 \text{ mm}, m_2 = 8.1 \text{ mm}, n_1 = 119.4 \text{ mm}, n_2 = 42.5 \text{ mm}.$$

The targets are numbered from a to d . And their ranging results are listed in Table 4.

According to Fig. 5(a) target a and b are in region **RB**, their values of $|E_x|$ are much smaller than those of target c and d . Target c is

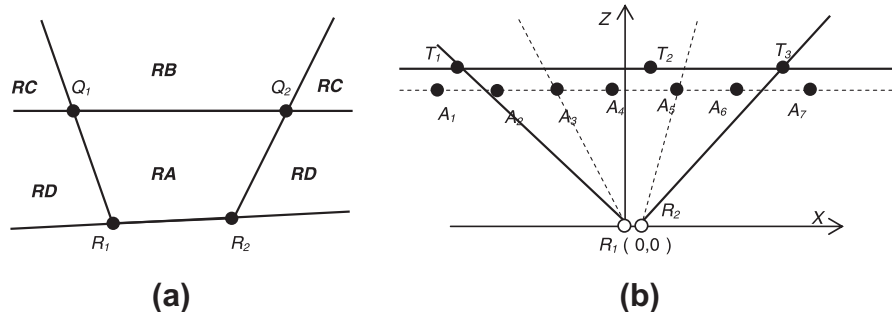


Fig. 5. The regions divided by R_1, R_2, Q_1 and Q_2 . (a) The relation between the target and the region determined by R_1, R_2, Q_1 and Q_2 . (b) The regions divided by R_1, R_2 and different points in Fig. 6.

Table 1
The information of spatial points in Fig. 5(b).

Point	A_1	A_2	A_3	A_4	A_5	A_6	A_7	T_1	T_2	T_3
(u_1, u_2)	(32,9)	(80,57)	(118,95)	(157,134)	(198,175)	(239,216)	(290,267)	(59,39)	(175,155)	(259,239)
x_{0t}/mm	−672.5	−406.5	−201.5	−1.5	199.9	398.5	688.5	−611.5	98.5	598.5
z_{0t}/mm	2016	2016	2016	2016	2016	2016	2016	2323	2323	2323

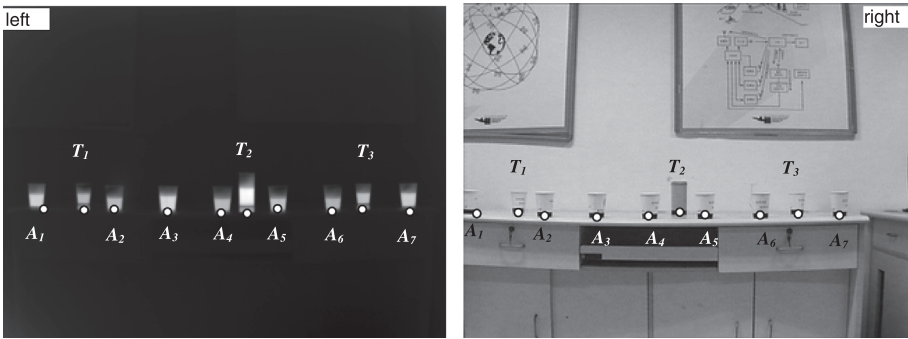


Fig. 6. Real images captured by multimodal stereo system in parallel configuration.

Table 2
Experimental results of different targets computed with different pairs of reference points in Fig. 6.

	(Q_1, Q_2)	(A_1, A_2)	(A_3, A_4)	(A_3, A_5)	(A_3, A_6)	(A_4, A_5)	(A_6, A_7)	(T_1, T_2)	(T_1, T_3)	(T_2, T_3)
Target										
T_1	region	RB	RC	RC	RC	RC	RC	/	/	/
	$ E_x /\text{mm}$	4.7	21.1	29.5	34.0	45.9	109.3	/	/	/
	$ E_z /\text{mm}$	57.8	30.6	23.3	19.5	16.4	67.4	/	/	/
T_2	region	RC	RC	RB	RB	RB	RC	/	/	/
	$ E_x /\text{mm}$	35.3	5.1	1.5	5.0	0.7	49.7	/	/	/
	$ E_z /\text{mm}$	57.8	30.6	23.3	19.5	16.4	67.4	/	/	/
T_3	region	RC	RC	RC	RC	RC	RB	/	/	/
	$ E_x /\text{mm}$	78.4	7.7	9.7	19.0	17.9	7.6	/	/	/
	$ E_z /\text{mm}$	57.8	30.6	23.3	19.5	16.4	67.4	/	/	/
A_2	region	/	/	/	/	/	/	RA	RA	RD
	$ E_x /\text{mm}$	/	/	/	/	/	/	6.5	8.8	7.6
	$ E_z /\text{mm}$	/	/	/	/	/	/	38.4	34.4	28.9
A_5	region	/	/	/	/	/	/	RD	RA	RA
	$ E_x /\text{mm}$	/	/	/	/	/	/	6.2	3.3	3.2
	$ E_z /\text{mm}$	/	/	/	/	/	/	38.4	34.4	28.9

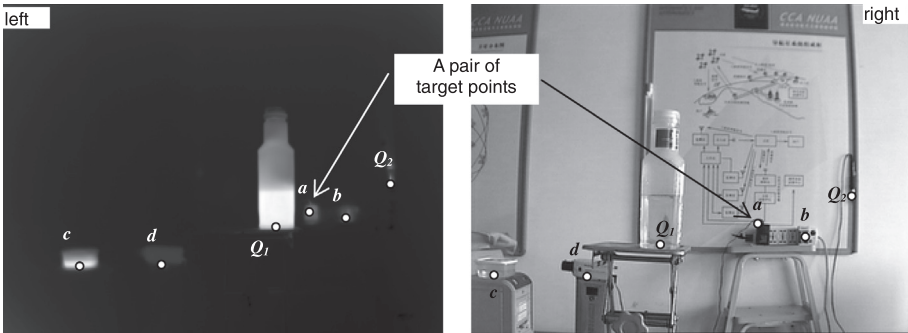


Fig. 7. Real images captured by another multimodal stereo system in parallel configuration.

Table 3
The information of two reference points in Fig. 7.

Q_i	l_1/mm	l_2/mm	X_0/mm	Z_{01}/mm	u_1	u_2
Q_1	948	829	77.5	945	199	139
Q_2	2533	2375	750	2419	281	277

neither in region **RB** nor close to the principal point. So its ranging results are the worst both in X-axis and Z-axis. Besides, the results of target *a* and target *d* show that when the targets are close to the principal point in the image plane, $|E_z|$ of the one close to the camera in Z-axis is smaller. One can infer that the longer the distance from the target to the camera in Z-axis is, the less pixels of

Table 4

Experimental results of different targets in Fig. 7.

target	(u_1, u_2)	x_{0t}/mm	z_{01t}/mm	$ E_x /\text{mm}$	$ E_z /\text{mm}$
a	(224,212)	356	2334	12.8	165
b	(250,242)	521	2349	1.2	55.4
c	(56,18)	-460	1590	70	198.7
d	(117,89)	-269	1890	21.6	9.6

the target there are on the image and the target's image on the thermal image plane is fainter. It makes the mismatch severer. So the error of the range computation is larger.

3.2. Experimental result of a moving target

In this section, the range from a moving person to the left rotation center of the pan-tilt device 2-D image sequences were captured by the multimodal stereo ranging system as the people moved throughout the indoor environment. Part of the image sequences are shown in Fig. 8. Fig. 8(a1) to (e1) are thermal images and Fig. 8(a2) to (e2) are visible images. The person is 178 cm in height. The room temperature was about 20 °C. As the head is hotter than the background, it is brighter in thermal images. The center of the head is selected as the target point and it is extracted through background subtraction and thresholding segmentation. In order to determine the accuracy of the range computed for tracking, the scene was calibrated too. This was done by having a person walk around the test-bed area, stopping at preset locations in the scene. At each location the range from the person's head to R_1 was measured to be used as ground truth. In this environment, one could get the values of these parameters after measurement: $r = 243$ mm, $t = 65$ mm. The range is computed based on the parameters identified in last section. The experimental results are listed in Table 5. The ranging results are quantified in terms of relative error E_l and E_θ in this part:

$$E_l = |l_{1t} - l_c| / l_{1t} \times 100\% \quad (24)$$

$$E_\theta = |\theta_t - \theta_c| / |\theta_t| \times 100\% \quad (25)$$

Here l_{1t} and θ_t are the ground-truth data, and l_c and θ_c are the computed results.

Both average relative errors of the line distances are less than 8% and both average relative errors of the angle are less than 1%.

The accuracy of the proposed method can basically meet the demand of object tracking. Table 5 shows that the ground truth l_{1t} of the person stopped at different locations had the relation as follows: $l_{1t}(e) < l_{1t}(a) < l_{1t}(d) < l_{1t}(b) < l_{1t}(c)$. And the results computed with the parameters identified by two reference points show that $E_l(e) < E_l(a) < E_l(d) < E_l(b) < E_l(c)$ and $E_\theta(e) < E_\theta(a) < E_\theta(d) < E_\theta(b) < E_\theta(c)$. That is to say as the distance between the person and the camera increases, the relative errors of the range and the angle are enlarged. As analyzed in Section 3.1, the longer the distance from the target to the camera is, the less pixels of the target there are on the image and the target's image on the image plane is fainter. So the error is larger. In addition, the corresponding points were matched manually so the corresponding points could not be matched precisely. The errors in this experiment were greatly caused by the mismatch. It will be improved in the future.

The distances from the reference points to R_1 were manually measured by a metric stick in the experiment. The accuracy of parameters calibration is greatly dependent on the measurement accuracy. More reference points can be used in calibration for robustness. The calibration error caused by the measurement error could be partly reduced by increasing the measurement time and then the ranging accuracy will be improved. If there are more than two reference points, Eq. (20) will be extended to overdetermined linear equations and optimized iterative algorithm must be used to search for the optimum solutions of m_1 , n_1 , m_2 and n_2 . Thus the computational cost will be increased. Three points in Fig. 7 are selected as the reference points to identify the unknown parameters in this experiment and the Levenberg–Marquardt (LM) method is used. The results are shown as follows: $m_1 = -33.3$ mm, $m_2 = 4.1$ mm, $n_1 = 83.7$ mm, $n_2 = 58.5$ mm.

Then the range between the moving person and the camera can be computed based on this unit of identified parameters. The results are shown in Table 5 too.

The results show that the accuracy is increased compared with the range computed with the parameters identified by two reference points. However, the accuracy is only slightly improved. Although the ranging accuracy of (b) to (d) is improved, the ranging accuracy of (a) and (e) is degraded. It is because that the measurement error caused by the measure method could not be eliminated by adding the measurement time. So it is very important to improve the measurement accuracy. It appears that two reference points are enough when great accuracy it not needed.

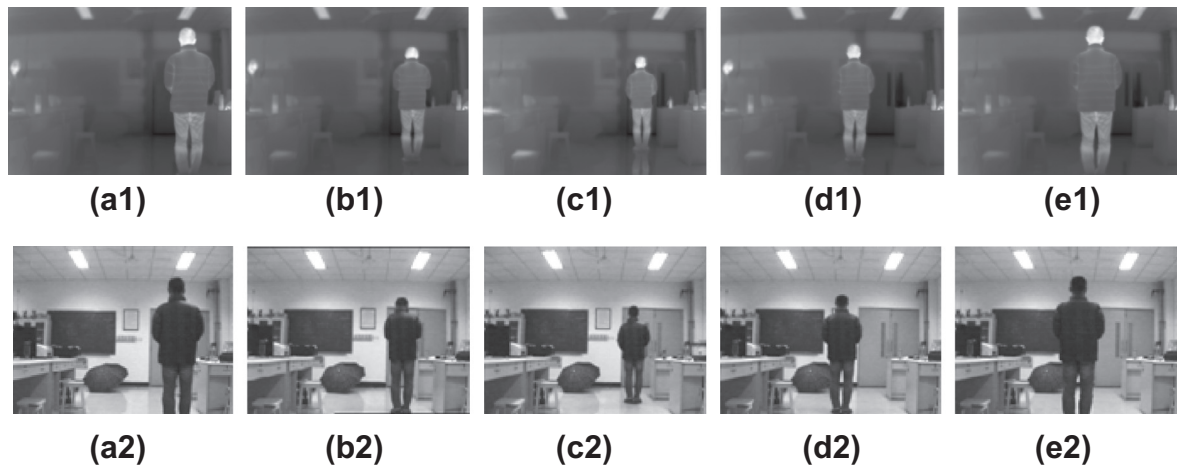


Fig. 8. The moving person captured by multimodal stereo system in parallel configuration. (a1) The thermal image captured at the first location. (b1) The thermal image captured at the second location. (c1) The thermal image captured at the third location. (d1) The thermal image captured at the fourth location. (e1) The thermal image captured at the fifth location. (a2) The visible image captured at the first location. (b2) The visible image captured at the second location. (c2) The visible image captured at the third location. (d2) The visible image captured at the fourth location. (e2) The visible image captured at the fifth location.

Table 5
Experimental results of the moving person.

No.	(u_1, u_2)	Ground truth		The results computed with the parameters identified by two points				The results computed with the parameters identified by three points			
		l_{1t}/mm	θ_t/rad	l_{1c}/mm	θ_c/rad	$E_{1l}/\%$	$E_{\theta}/\%$	l_{1c}/mm	θ_c/rad	$E_{1l}/\%$	$E_{\theta}/\%$
a	(256,235)	2640.3	1.340	2764.5	1.332	4.70	0.58	2893.3	1.337	9.58	0.21
b	(238,221)	3423.9	1.393	3068.7	1.377	10.37	1.15	3186.9	1.381	6.92	0.87
c	(225,214)	4213.7	1.427	3709.9	1.410	11.96	1.18	3832.7	1.413	9.04	0.98
d	(191,174)	3375.9	1.512	3035.2	1.499	10.09	0.86	3083.7	1.501	8.65	0.69
e	(196,174)	2577.8	1.493	2643.4	1.486	2.55	0.46	2690.2	1.489	4.36	0.25
Average error	/	/	/	/	/	7.93	0.85	/	/	7.71	0.6

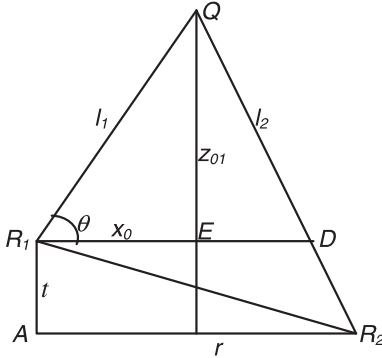


Fig. 9. The geometric relationship of the rotation centers and the reference point Q.

In addition, only linear model is discussed in the proposed method. The lens distortions are ignored here. The principle point of the thermal camera is not precisely calibrated. And there is halation in thermal images. All these factors mentioned above have contributed to the calibration error, therefore increasing the ranging errors.

A ranging system composed of a thermal camera and a visible camera arranged in parallel configuration is built in the experiment. It is used to compute the range between different target points and the camera in different scenes by the proposed method. It can be easily operated. For a certain target, the effect of the relative position of the target and reference points is analyzed when only two reference points are used to identify the parameters. The subsequent experiments are carried out based on the analysis. Furthermore, the calibrated stereo system can be used in the scene that is different from the calibration scene. The average relative error of the range computation with the proposed method is less than 8%. Considering that most of the distance parameters are manually measured in the proposed method and the thermal images can not be as clear as the visible images, the performance is acceptable for object tracking.

4. Conclusions

As the thermal camera is different from the visible camera, a ranging method based on the rotation center of the pan-tilt device was proposed in this paper. The range can be computed by the proposed method in a stereo system composed of a thermal camera and a visible camera arranged in parallel configuration. It will contribute to precise object tracking and locating in all weather conditions. In addition, a simple, fast and reliable method for determining the parameters required for range computation proposed in this paper was presented too. The information of the reference points used in calibration can be measured easily. The method is able to be easily implemented with a pair of visible-thermal cameras on most computer systems at a very low cost

and in an efficient way. Experiments with real images proved the validity of the method.

Acknowledgements

The work is subsidized by the National Natural Science Foundation of China (Grant No. 81071236) and Jiangsu Prospective Joint Research Project (BY2013003-12).

Appendix A

In practical operation, r , t can be measured directly. However, the measurement of x_{01} and z_{01} is more complicated, and bigger errors might be made during the measurement. By contrast, the line distances from Q_1 (Q_2) to each rotation center (l_1 and l_2) are easier to be measured. As shown in Fig. 9, the geometric relationship between the line distances and x_{01} or z_{01} is certain. So x_0 and z_{01} can be figured out by Eqs. (26) and (27).

$$Z_{01} = l_1 \sin \left(\arccos \left(\frac{l_1^2 + t^2 + r^2 - l_2^2}{2l_1 \cdot \sqrt{t^2 + r^2}} \right) - \arctan \left(\frac{t}{r} \right) \right) \quad (26)$$

$$X_0 = l_1 \cos \left(\arccos \left(\frac{l_1^2 + t^2 + r^2 - l_2^2}{2l_1 \cdot \sqrt{t^2 + r^2}} \right) - \arctan \left(\frac{t}{r} \right) \right) \quad (27)$$

Let AR_1 be the vertical distance between R_1 and R_2 . Let AR_2 be the horizontal distance between R_1 and R_2 . $AR_1 = t$ (signed) and $AR_2 = r$. When R_2 is behind R_1 , $t > 0$. Let QE be the vertical distance between R_1 and Q . Let R_1E be the horizontal distance between R_1 and Q . $QE = z_{01}$ and $R_1E = x_0$. Let QR_1 be the line distance between R_1 and Q . Let QR_2 be the line distance between R_2 and Q . $QR_1 = l_1$ and $QR_2 = l_2$. The parameters below are to be worked out: $QE = z_{01}$, $R_1E = x_0$.

According to the Pythagorean Theorem one can get: $R_1R_2 = AR_1^2 + AR_2^2 = t^2 + r^2$

$$\therefore \tan \angle R_1R_2A = AR_1/AR_2 = t/r$$

$$\therefore \angle DR_1R_2 = \angle R_1R_2A = \arctan \left(\frac{t}{r} \right)$$

$$\begin{aligned} \text{According to the law of cosines: } \cos \angle QR_1R_2 &= \frac{QR_1^2 + R_1R_2^2 - QR_2^2}{2QR_1 \cdot R_1R_2} \\ &= \frac{l_1^2 + t^2 + r^2 - l_2^2}{2l_1 \cdot \sqrt{t^2 + r^2}} \end{aligned}$$

$$\angle QR_1R_2 = \arccos \left(\frac{l_1^2 + t^2 + r^2 - l_2^2}{2l_1 \cdot \sqrt{t^2 + r^2}} \right)$$

$$\begin{aligned}
\angle QR_1 D &= \angle QR_1 R_2 - \angle DR_1 R_2 \\
&= \arccos \left(\frac{l_1^2 + t^2 + r^2 - l_2^2}{2l_1 \cdot \sqrt{t^2 + r^2}} \right) - \arctan \left(\frac{t}{r} \right) \\
z_{01} &= QE = QR_1 \sin \angle QR_1 D \\
&= l_1 \sin \left(\arccos \left(\frac{l_1^2 + t^2 + r^2 - l_2^2}{2l_1 \cdot \sqrt{t^2 + r^2}} \right) - \arctan \left(\frac{t}{r} \right) \right) \\
x_0 &= R_1 E = QR_1 \cos \angle QR_1 D \\
&= l_1 \cos \left(\arccos \left(\frac{l_1^2 + t^2 + r^2 - l_2^2}{2l_1 \cdot \sqrt{t^2 + r^2}} \right) - \arctan \left(\frac{t}{r} \right) \right)
\end{aligned}$$

References

- [1] Wang Leilei, Target range based on visual and infrared automotive night vision system[D], Zhejiang: Zhejiang Sci-Tech University, 2011 (in Chinese).
- [2] G. Pajares, J.M. de la Cruz, A wavelet-based image fusion tutorial [J], *Pattern Recogn.* 37 (9) (2004) 1855–1872.
- [3] S.G. Kong, J. Heo, B.R. Abidi, et al., Recent advances in visual and infrared face recognition – a review[J], *Comput. Vis. Image Underst.* 97 (1) (2005) 103–135.
- [4] S.J. Krotosky, M.M. Trivedi, Person surveillance using visual and infrared imagery [J], *IEEE Trans. Circuits Syst. Video Technol.* 18 (8) (2008) 1096–1105.
- [5] M. Lauer, M. Schönbein, S. Lange, et al., 3D-objecttracking with a mixed omnidirectional stereo camera system [J], *Mechatronics* 21 (2) (2011) 390–398.
- [6] Y. Satoh, T. Okatani, K. Deguchi, Binocular motion tracking by gaze fixation control and three-dimensional shape reconstruction [J], *Advanced Robotics* 17 (10) (2003) 1057–1072.
- [7] R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision* [M], Cambridge University Press, Cambridge, UK, 2006.
- [8] R.Y. Tsai, A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses [J], *IEEE J. Rob. Autom.* 3 (4) (1987) 323–344.
- [9] Zhang Zhengyou, A flexible new technique for camera calibration [J], *IEEE Trans Patten Anal. Machine Intell.* 22 (11) (2000) 1330–1334.
- [10] X.Q. Meng, Z.Y. Hu, A new easy camera calibration technique based on circular points [J], *Pattern Recogn.* 36 (5) (2003) 1155–1164.
- [11] S. Qjucheng, D. Xiaogang, L. Ming, et al., An improved method of camera calibration [C], *Electronic Measurement Instruments (ICEMI)*, 10th International Conference on, IEEE, 2011, pp. 155–157.
- [12] P.J. Da Silva Tavares, M.A. Vaz, Accurate subpixel corner detection on planar camera calibration targets [J], *Opt. Eng.* 46 (10) (2007).
- [13] S.J. Krotosky, M.M. Trivedi, Mutual information based registration of multimodal stereo videos for person tracking [J], *Comput. Vis. Image Underst.* 106 (2) (2007) 270–287.
- [14] X. Ju, J. Nebel, J.P. Siebert, 3D thermography imaging standardization technique for inflammation diagnosis [J], *Proc. Spie, Photonics Asia* 5640 (2004) 266–273.